DOCUMENT RESUME

ED 211 081 IR 009 903

AUTHOR	Craven, Timothy C.; Fjerestad, Li
TITLE	NEPHIS Implementation Title Coding Manual. Third Edition.
INSTITUTION	University of Western Ontario, London. School of Library and Information Science.
SPONS AGENCY	Natural Sciences and Engineering Research Council, Ottawa (Ontario).
PUB DATE	81
GRANT	A3027
NOTE	25p.
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Autoinstructional Aids; *Automatic Indexing; Guidelines; Indexing; Information Science; *Permuted Indexes; Program Implementation; *Subject Index Terms
IDENTIFIERS	*Coding; *NEsted PHrase Indexing System

ABSTRACT

Prepared as a self-instructional tool for anyone interested in indexing documents from their titles using the NEsted PHrase Indexing System (NEPHIS), this manual introduces, defines, and provides guidelines for the use of this computer-assisted indexing system. The objectives of the system are outlined, the NEPHIS computer program and input formats are discussed, and a detailed set of NEPHIS implementation rules covers such topics as the standardization of titles, potential errors, and access points. A NEPHIS implementation sheet, a description of a project in which NEPHIS was implemented on a set of over 3,000 Canadian government documents, and a list of bibliographic references are appended. A NEPHIS index sample page is also attached. (JL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED211081

NEPHIS IMPLEMENTATION TITLE CODING MANUAL

THIRD EDITION

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Timothy C. Craven

and

Li Fjerestad

School of Library and Information Science

The University of Western Ontario

London, Ontario

N6A 5B9

1 9 8 1

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

T.C. Craven

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

IR009903

TABLE OF CONTENTS

Introduction 3

- Purpose of the Title Coding Manual 3
- What is NEPHIS 3
- Objectives of NEPHIS 4

Subject Indexing and Title Coding 5

Simplified NEPHIS Format - Nesting of Phrases 6

- The "<" and ">" Brackets 6

Expanded NEPHIS Format 8

- The "@" Sign 8
- The "?" Mark 9
  - As Forward-Reading Connective Tag 9
  - As Backward-Reading Connective Tag 10

NEPHIS Implementation Rules 12

- Preliminary Rules 12
  - Dealing with Unusual Characters in Titles 12
  - Standardization of Titles 12
  - Fatal Error Conditions 14
  - Error Conditions in Character Spacing 15
- Rules Governing Access Points 16
- Nesting Procedure Rules 17
  - General Nesting Rules 17
  - Forward-Reading Connective Rules 18
  - Backward-Reading Connective Rules 19

Appendix 1 Implementation Reference Sheet 22

Appendix 2 Implementation Project History 23

Appendix 3 Bibliographic References 24

Figure 1 NEPHIS Index Sample Page 25

INTRODUCTION

Purpose of the Title Coding Manual

This brief manual of practice has been prepared as a self-instructing tool for anyone interested in indexing documents from title using the NEPHIS system. Readers may be beginning or experienced indexers, NEPHIS coders, or researchers interested in further work on the NEPHIS system. Previous training in library science or indexing has not been assumed, particularly in the beginning sections.

The NEPHIS indexing system will be introduced, subject indexing and indexing from title will be mentioned, and NEPHIS indexing rules which have been developed will be presented. The Title Coding Manual originally contributed to the documentation of a two-year project: Implementation of NEPHIS on a Set of Scientific/Technical Documents. See Appendix 2 for information on project history, indexing data elements and material indexed.

What is NEPHIS?

NEPHIS (NEsted PHrase Indexing System) is a system of computer-assisted indexing. Basically the NEPHIS program takes a single input statement worked out by the human indexer and produces from it a whole set of entries for the index.

The NEPHIS computer program notes the access terms which the human indexer has marked by means of the NEPHIS symbols. The program reads each input string sequentially and produces the appropriate permutations (index entries), which are all complete representations of the original subject or title. See Appendix 1 for examples of input strings and permutations.

The output strings are sorted and formatted by the computer. This sorted and formatted file of permutations results in a printed or otherwise visually-displayed index which is elegant and browsable. Figure 1 is a sample from a NEPHIS index. Notice that no special coding symbols appear in the finished index.

Objectives of NEPHIS

The system was designed with a number of fairly specific objectives in mind. Simplicity of the system is basic to its objectives. The system should be easy for the indexer, easy to program, inexpensive to run and easy for the user of the index.

(1) Easy for the Indexer. The system should be easy for the naive indexer to learn, and once learned, it should be easy to apply. This requirement implies certain features which have been included in NEPHIS: closeness of the input to natural language, a minimum number of additional features to be learned and used consistently, and shortness of input.

(2) Easy to Program. The algorithm that constitutes the computer-part of the NEPHIS system should be fairly simple, and possible to implement in a short time on virtually any suitable machine in any programming language capable of string manipulation. Although NEPHIS has been developed as a computer-implemented system, its simplicity also allows the option of a manually-produced NEPHIS index.

(3) Inexpensive to Run. Certain features help meet this objective: not requiring large random-access files or occupying a large amount of memory, and the relatively simple algorithm which cuts down on both core requirements and processor time.

(4) Easy for the user of the index. The fundamental requirement for any system is that the index produced should be usable. The output must be: reasonably elegant in format, fairly close to natural language, and easy to browse through. Finally, any special symbols used in the finished index should be kept to a minimum and be consistent with common indexing practice.

SUBJECT INDEXING AND TITLE CODING

The human indexer prepares an input file of records by basically one of two methods: indexing by assignment or indexing from title. Although this manual is concerned with indexing from title, indexing by assignment is mentioned for general background information.

In indexing by assignment, the human indexer decides on suitable document descriptions and encodes these by means of indexing symbols into linear sequences of words and symbols called input strings. (Examples of NEPHIS input strings are found in Appendix 1 and later in the text of this manual.)

The NEPHIS indexer is generally advised in indexing by assignment to use nouns and propositions as much as possible in the construction of input strings. This feature of NEPHIS is based on the fact that prepositions in indexing languages regularly function as "constituent delimiters". Controlled vocabulary may be found situationally advisable to handle both spelling variations and synonyms. It may also be desirable, for example, to suppress broad index access terms which might be followed by pages of index entries.

This manual is concerned with the second indexing approach - title coding. In indexing from title, the original input strings are supplied by the document titles themselves with special indexing symbols added. The title is coded, but remains basically unchanged.

The quality of the resulting index therefore depends directly on the precision and descriptive quality of the original titles. The process of coding titles is not suitable when the titles inadequately reflect the subject content. However, the titles of government documents for example (especially in the scientific and technical fields) are often fairly thorough descriptions of content. In such cases the title coding approach may be both suitable and less expensive.

SIMPLIFIED FORMAT - NESTING OF PHRASES

The "<" and ">" Brackets

NEPHIS input strings are in many ways similar to natural language. For example:

PALEOZOIC <GEOLOGY OF THE <BRUCE PENINSULA AREA, <ONTARIO>>>.

Usually the input string takes the basic form of a noun phrase with other noun phrases nested (or embedded) within it. An indexer needs only two coding symbols to nest phrases. These tagging characters are the left angular bracket "<" and the right angular bracket ">". The simplified NEPHIS format is based on only these two characters. The following is another typical input string using the simplified format:

PETROLEUM <REFINERIES IN <CANADA>>.

This statement instructs the computer program to generate these permutations:

PETROLEUM REFINERIES IN CANADA.

REFINERIES IN CANADA. PETROLEUM -.

CANADA. REFINERIES IN -. PETROLEUM -.

If the input string produced by the indexer read:

PETROLEUM <REFINERIES IN CANADA>.

The NEPHIS program would generate these permutations:

PETROLEUM REFINERIES IN CANADA.

REFINERIES IN CANADA. PETROLEUM -.

The underlying mechanism is recursive. That is, the indexer may indicate that a phrase is nested within another phrase which itself may be nested within a third or fourth phrase. The order of the elements in one or more of the permutations may be improved and the readability of the entries increased by proper use of this mechanism.

Normally there will be one entry for the whole input string and an additional entry for each nested phrase.

Words or phrases appearing only once in the title are not repeated more than once in any permutation. Therefore the point where a phrase that has already appeared in the permutation has been omitted from its position in natural language sequence is indicated by a space-plus-dash:

REFINERIES IN CANADA. PETROLEUM -.

The NEPHIS program always deals with the larger nested phrases in which the lead phrase is embedded before dealing with the rest of the input string.

Notice that the tagging characters disappear in the permutations produced by the computer. Notice also that the beginning of an input string does not require a left angular bracket.

The human indexer should keep in mind that usually in the NEPHIS simplified format only the phrases which are desired as lead index terms should be nested.

There must in all cases be the same number of left and right angular brackets in an input string. Unmatched brackets cause a fatal error situation, and no permutations will be formed from an input string in which brackets are unmatched. Notice that the left angular bracket directly precedes a nested phrase, and that a right angular bracket directly follows a nested phrase, with no intervening spaces.

In general, blank spaces and punctuation marks have significance in computer sorting and should therefore be checked carefully to insure consistency in, and appropriate grouping of, index entries.

EXPANDED NEPHIS FORMAT

Three other features have been added to the NEPHIS simplified format to increase precision and readability in the index entries. These features are represented by the "@", the "?" tagging a forward-reading connective, and the "?" tagging a backward-reading connective. (See Appendix 1 for short explanations.)

THE "@" SIGN: Suppression of Undesired Points of Access

The number of permutations produced by the computer from an input string can usually be calculated by counting one permutation for the whole input string and an additional entry for every nested phrase. Some initial words (VARIATIONS or MEASUREMENT, for example) are relatively useless as points of access in a printed index. In such cases the "@" sign may be used to suppress an access point which is not desired. The "@" appears directly before the term which is to be omitted as an access point.

@CATALOGUE OF <AIRPHOTOGRAPHS>.

AIRPHOTOGRAPHS. CATALOGUE OF -.

@HOW TO ORDER A <TOPOGRAPHICAL <MAP>>.

TOPOGRAPHICAL MAP. HOW TO ORDER A -.

MAP. TOPOGRAPHICAL -. HOW TO ORDER A -.

By means of the "@" sign only the indexer-chosen permutations appear in the printed index.

The "@" can also be used to suppress undesired points of access located at the beginnings of nested phrases. Potential permutations tagged by an "@" sign will not appear in the printed index.

GEOLOGY OF <@SOUTHEASTERN <PRINCE EDWARD ISLAND>>.

GEOLOGY OF SOUTHEASTERN PRINCE EDWARD ISLAND.

PRINCE EDWARD ISLAND. SOUTHEASTERN -. GEOLOGY OF -.

The "@" sign may be employed at the beginnings of both the input string and nested phrases if so desired by the indexer.

@IDENTIFICATION OF <@ANCIENT <CRATERS>>.

CRATERS. ANCIENT -. IDENTIFICATION OF -.

THE "?" MARK

The other two special features of NEPHIS were added to the system to improve readability and intelligibility of the permutations. Both involve the use of the question mark to indicate that the following characters form a connective element between words or phrases.

"?": As Forward-Reading Connective Tag

A forward-reading connective always precedes the nested phrase with which it is associated. It is introduced by a question mark which comes directly before it. A single left angular bracket indicates both the end of the forward-reading connective and the beginning of the nested phrase.

GEOLOGY? OF THE <NEW QUEBEC <CRATER>>.

GEOLOGY OF THE NEW QUEBEC CRATER.

NEW QUEBEC CRATER. GEOLOGY.

CRATER. NEW QUEBEC -. GEOLOGY.

In the example, the question mark has suppressed the phrase " OF THE " in two permutations. The same input string could also be coded using the simplified format, but the results shown below are less like natural language and not as readable. Notice that the phrase OF THE appears in

all permutations if the simplified format is employed.

GEOLOGY OF THE <NEW QUEBEC <CRATER>>.

GEOLOGY OF THE NEW QUEBEC CRATER.

NEW QUEBEC CRATER. GEOLOGY OF THE -.

CRATER. NEW QUEBEC -. GEOLOGY OF THE -.

A question mark can also be used to tag punctuation which is functioning as a connective element in an input string.

METALLIC <ORES>?: <MILLING PLANTS IN CANADA>.

METALLIC ORES: MILLING PLANTS IN CANADA.

ORES. METALLIC -: MILLING PLANTS IN CANADA.

MILLING PLANTS IN CANADA. METALLIC ORES.

"?": As Backward-Reading Connective Tag

A backward-reading connective always appears at the end of the nested phrase with which it is associated. Like the forward-reading connective, it is introduced by a question mark. A single right angular bracket marks the end of the connective and of the nested phrase. The most common purpose of the backward-reading connective is to deal with coordinating expressions such as AND or OR so that the vital connective will appear in each permutation. The question mark should be placed directly before the first character of the backward-reading connective. Notice that the connective appears in each permutation only once, and that in all cases there is an equal number of left and right angular brackets in the input string.

JAMES BAY? AND <HUDSON BAY? AND > <DRAINAGE>.

JAMES BAY AND HUDSON BAY DRAINAGE.

HUDSON BAY AND JAMES BAY DRAINAGE.

DRAINAGE. JAMES BAY AND HUDSON BAY -.

If the forward-reading connective were marked without the backward-reading connective being present the results would have been unacceptable. Notice in the example below that the second permutation is unintelligible.

JAMES BAY? AND <HUDSON BAY> <DRAINAGE>.

JAMES BAY AND HUDSON BAY DRAINAGE.

HUDSON BAY. JAMES BAY DRAINAGE.

DRAINAGE. JAMES BAY AND HUDSON BAY -.

Finally, if the simplified format is used for the same input string, the second permutation again lacks readability.

JAMES BAY AND <HUDSON BAY> <DRAINAGE>.

JAMES BAY AND HUDSON BAY DRAINAGE.

HUDSON BAY. JAMES BAY AND - DRAINAGE.

DRAINAGE. JAMES BAY AND HUDSON BAY -.

The example above demonstrates the effectiveness of the backward-reading connective in improving the readability of certain permutations.

Notice that the addition of a backward-reading connective to a NEPHIS input string forms an exception to the rule that only NEPHIS coding characters are to be added to the title in title-based indexing. Because of the increase in clarity in some of the permutations, a backward-reading connective is added to an input string whenever the human indexer considers it necessary.

The backward-reading connective was introduced to deal with a variety of other problems by means of a single common mechanism. However, only this one function of continuity concerning coordinating elements has been found important in this implementation. Other uses of the backward-reading connective are discussed in reference 2.

NEPHIS IMPLEMENTATION RULES

PRELIMINARY RULES

Dealing with Unusual Characters in Titles

Characters which are not available on the printer or which may be confusing to the NEPHIS computer program must be dealt with in special ways which are basically of three types: omission, naming and substitution.

1. Certain characters should be omitted in the coding process if they occur in the original titles. Omit:

   question marks,

   quotation marks,

   diacritic marks.

2. Most characters should be named if they appear in the original titles as symbols. A character should be spelled out if its name would be used in natural spoken language.

   Spell out the symbol for "degree" as "degree" or "degrees"

   and the symbol for "percent" as "percent".

   Name any Greek letters, as in:

   ALPHA ENERGY SPECTROMETRY.

3. The following substitutions should be used:

   For "[" substitute "(".

   For "]" substitute ")".

Standardization of Titles

The integrity of the title is basic to the NEPHIS title coding system in general and to this implementation in particular. However certain rules were developed to standardize parallel title structures in order to promote consistency in the indexing results.

1. Do not begin with a space.

2. Omit the initial articles A, AN, and THE. Retain any articles in other positions.

3. Follow punctuation marks such as commas, periods or semi-colons with single spaces.

4. Always end with a period.

5. Maintain the internal punctuation of a title in the coding process if title meaning or integrity of structure is dependent upon it.

6. Omit periods in abbreviations.

7. Use a hyphen in compound words or place names consisting of only two words. If the compound word or place name consists of three or more words, use a space-hyphen-space to separate the elements.

8. In cases where certain punctuation is not vital to integrity and where random differences in parallel titles already exist, change punctuation to promote consistency in the printed index entries. For example, make punctuation in series publications consistent with the general existing pattern.

9. In making changes to promote punctuation consistency, do not exceed reasonable limits of time and cost-effectiveness.

Here are some examples which demonstrate the correct procedural rules for the standardization of titles and punctuation in this implementation.

GAZETTEER OF CANADA: MANITOBA.

22.25 MHZ RADIO TELESCOPE.

VOL 7, NOS 1, 2, 3 AND 4, 1965.
VOL 8, NO 3, 1966.
VOL 8, NO 4, 1966.

GEOLOGY OF IRON DEPOSITS IN
CANADA. VOLUME 2: IRON DEPOSITS.

ST LAWRENCE PILOT, 1963 EDITION.
BRITISH COLUMBIA PILOT, 1966 EDITION.

HALIFAX - SABLE ISLAND MAP-AREA.

STRESS-CORROSION CRACKING TESTS ON SOME
HIGH-STRENGTH STEELS.

Fatal Error Conditions

In some cases errors in title coding are made which prohibit the formation of any permutations from an output string. These errors are known as fatal errors and there are four types in this implementation.

1. Match each right angular bracket with a left angular bracket. In any input string there must be an equal number of right and left angular brackets.

   CANADIAN <IRON ORE <INDUSTRY>. (Fatal Error)

   CANADIAN <IRON ORE INDUSTRY>>. (Fatal Error)

   CANADIAN <IRON ORE <INDUSTRY>>. (Correct)

2. End each input string with a period. This rule is not a general rule of the NEPHIS system, but failure to adhere to it does constitute a fatal error in this implementation. No permutations will result from an input string which does not end in a period.

   CERAMIC PLANTS? IN <CANADA> (Fatal Error)

   CERAMIC PLANTS? IN <CANADA>. (Correct)

3. Do not include a tagging character within a connective. If duplicate question marks appear in series without being separated by a right or left angular bracket, a fatal error occurs and no permutations will be produced.

   STABILITY? OF SLOPES? IN <OPEN PITS>. (Fatal Error)

   STABILITY? OF <SLOPES? IN <OPEN PITS>>. (Correct)

4. Do not use a connective tag not followed by a left or right angular bracket.

   CANADIAN <MINERALS YEARBOOK>?, 1965. (Fatal Error)

   CANADIAN <MINERALS YEARBOOK>, 1965. (Correct)

Error Conditions in Character Spacing

Most errors are nonfatal: that is, they allow the formation of permutations but the permutations are in an incorrect form. A common cause of such errors is faulty spacing of characters - resulting in words which are run together or extra spaces which cause problems in computer sorting. Spacing should be carefully checked by the beginning NEPHIS indexer.

1. Always directly precede a nested phrase with a left angular bracket. Spacing normally found in natural language is maintained, and the coding characters are simply added to the natural language structure.

   CANADIAN <MINERALS YEARBOOK>. (Correct)

   CANADIAN< MINERALS YEARBOOK>. (Error)

2. Always directly follow a nested phrase with a right angular bracket. If there is more than one right angular bracket following a nested phrase because more than one nested phrase ends at the same point, the second bracket should follow the first directly, etc.

   CANADIAN <IRON ORE <INDUSTRY>>. (Correct)

   CANADIAN <IRON ORE <INDUSTRY> >. (Error)

   CANADIAN <IRON ORE <INDUSTRY >>. (Error)

3. When suppressing an entry, insert the "@" sign directly before the initial word of the input string or the nested phrase to which it refers. No space should occur between an "@" sign and the undesired lead term, nor between an "@" sign and a left angular bracket beginning a nested phrase.

   @RESEARCH? IN <@IMPROVED METHODS? OF <ROCKS <BREAKAGE>>>. (Correct)

   @ RESEARCH? IN <@ IMPROVED METHODS? OF <ROCKS <BREAKAGE>>>. (Error)

4. Insert a question mark indicating a forward-reading connective before the connective, which may include an initial space. The tagging character will directly follow the word before the connective.

   STANDARD? FOR <TRIASSIC <TIME>>.

   CLEARWATER COMPLEX?, <NEW QUEBEC>.

5A. If a connective is a word, ensure that there is a space both before and after the word.

SAND? AND <GRAVEL? AND >. (Correct)

SAND? AND <GRAVEL? AND>. (Error)

5B. If a forward-reading connective ends in a left parenthesis, make the corresponding backward-reading connective a space plus a left parenthesis - " (" - and mark the end of the nested phrase immediately before the right parenthesis.

BRACHIOPODS? OF THE <DETROIT RIVER GROUP>?
(<DEVONIAN? (>)? FROM <@SOUTHWESTERN
<ONTARIO> AND ADJACENT AREAS OF <MICHIGAN?
AND <OHIO? AND >>>.

RULES GOVERNING ACCESS POINTS

1A. Provide access only under terms which indicate the subject of the document.

ALBERTA?, <WATER YEAR 1967>.

POLYMORPHISM? OF THE <OXIDE <TI3 O5>>.

1B. Do not provide access under terms considered unimportant or useless as entry points.

@CONTRIBUTIONS? TO <CANADIAN
<PALEONOTOLOGY>>.

@THREE MEASUREMENTS? OF <HEAT FLOW? IN
<@EASTERN <CANADA>>>.

1C. Do not provide access under terms indicating the form or type of the document. Included in this category are GAZETTEER, ANNUAL REPORT, BIBLIOGRAPHY, INDEX, etc.

@GAZETTEER? OF <CANADA>?: <SASKATCHEWAN>.

@THESAURUS? OF <MINING TERMS>.

2. If a word desired as an access point is used more than once in a title, use only one occurrence as an access point.

ATLANTIC COAST: VOLUME 1?, <@ATLANTIC COAST?
AND <BAY OF <FUNDY>? AND >>.

3. Use numbers as access points if they are important to the subject content of the title.

   @LARGE <10 MHZ <ARRAY? FOR <RADIO <ASTRONOMY>>>.

   @AUTOMATIC <3-MAGNET? OR <BIASTATIC? OR > <MAGNETOMETER>>.

4. If there are no access points at all which you consider appropriate in a title, use an asterisk "*" instead of the initial "@" sign.

   *HIGHLIGHTS '68.

   *VOL 12, INDEX.

NESTING PROCEDURE RULES

General Nesting Rules

1A. When a noun which is not part of a proper name is modified, without intervening punctuation, by both a prepositional phrase and by an adjective, group the noun first with the prepositional phrase and then with the adjective.

   PETROLEUM <REFINERIES? IN <CANADA>>.

   @FACTORS INFLUENCING THE <MINERAL <ECONOMY? OF <CANADA>>>: PAST, PRESENT AND FUTURE.

1B. Do not apply rule 1A in the case of prepositional phrases (or noun phrases) directly preceded by commas or other punctuation marks.

   MAGNETIC <SUBSTORMS>, DECEMBER 5, 1968.

   CRUSTAL <REFRACTION EXPERIMENT>?: <YELLOWKNIFE 1966>.

2. When a noun is modified by two adjectives, group the noun first with the second adjective and then with the first.

   @EVALUATION? OF <CANADIAN <COMMERCIAL <COALS>>>.

   SEISMOLOGICAL DETECTION AND IDENTIFICATION? OF <UNDERGROUND <NUCLEAR <EXPLOSIONS>>>.

3A. Nest within proper names containing desirable entry points.

JASPER <NATIONAL <PARK>> BEHIND MOUNTAINS AND GLACIERS. (Correct)

3B. Do not mark off a nested phrase consisting of part of a proper name plus characters which are not part of the proper name.

JASPER <NATIONAL <PARK BEHIND MOUNTAINS AND GLACIERS>>. (Error)

4. In this implementation, nest only whole words, not parts of words. This rule also applies to compound, hyphenated words.

@GROWTH? OF <ZINC-SULPHIDE CRYSTALS>. (Correct)

@GROWTH? OF <ZINC-<SULPHIDE> CRYSTALS>. (Error)

RUBIDIUM-STRONTIUM <ISOCHRON AGE STUDIES>, REPORT 1. (Correct)

Forward-Reading Connective Rules

1. Include an article as part of a forward-reading connective if the article is followed by a nested phrase.

@UPPER <PLIENSBACHIAN> BEDS? IN THE <FERNIE GROUP? OF <ALBERTA>>.

ANORTHOSITE? OF <@NORTHERN <CAPE BRETON ISLAND?, <NOVA SCOTIA>>>?, A <PETROLOGICAL ENIGMA>.

2A. If there is one volume or part number described in a title, mark the beginning of a forward-reading connective directly after the designation of the volume or part number.

@VOLUME 2?: <GULF OF <ST LAWRENCE>>.

2B. If the title describes more than one volume or part, do not mark forward-reading connectives after the volume or part designations.

PACIFIC COAST: VOLUME 5, <JUAN DE FUCA? AND <GEORGIA? AND > STRAITS>; VOLUME 6, <BARKLEY SOUND>.

3. Do not mark a connecting phrase with a forward-reading connective tag if it is considered necessary for the readability of all permutations.

   GEOLOGY? OF THE <PORT RADIUM MAP-AREA?, <DISTRICT OF <MACKENZIE>>>.

   DEPARTMENT OF <ENERGY, MINES AND RESOURCES>, ANNUAL REPORT 1975/1976.

   BRITISH COLUMBIA PILOT, VOLUME 1?: <@SOUTHERN PORTION? OF THE <COAST OF BRITISH COLUMBIA>? INCLUDING THE <COAST OF THE <UNITED STATES> AND OF BRITISH COLUMBIA>>.

   (Note in the last example that an index entry beginning "UNITED STATES. COAST OF THE - AND OF BRITISH COLUMBIA. ..." would make sense, whereas one beginning "UNITED STATES. COAST AND OF BRITISH COLUMBIA. ..." would not.)

Backward-Reading Connective Rules

1. Use the backward-reading connective generally in this implementation only for coordination - when the elements of the title are coordinated by AND, by OR, or by parentheses.

   ROCK? AND <MINERAL? AND > COLLECTING? IN <CANADA>.

2. Do not use the coordinating word AND as a backward-reading connective if some of the resulting permutations would not accurately reflect the original meaning. Expressions like AND ITS, for example, should be encoded using the simplified format.

   GEOCHEMISTRY? OF <SILVER AND ITS <DEPOSITS>>. (Correct)

   GEOCHEMISTRY? OF <SILVER? AND ITS <DEPOSITS? AND >>. (Error)

   (Note how the index entry "DEPOSITS. SILVER AND ITS -. GEOCHEMISTRY." accurately reflects the relationship between SILVER and DEPOSITS in this title, whereas "DEPOSITS AND SILVER. GEOCHEMISTRY." would not.)

3. Use the word AND as a backward-reading connective when more than two elements are coordinated, the last two by AND

FLEXURAL RIGIDITY, THICKNESS AND VISCOSITY OF THE LITHOSPHERE. (Original Title)

S1, S2 AND S3 (Order of Elements in Original Title)

FLEXURAL <RIGIDITY>?, <THICKNESS? AND >? AND <VISCOSITY? AND >? OF THE <LITHOSPHERE>. (Encoded Title)

S1?, <S2? AND >? AND <S3? AND > (Order of Elements in Encoded Title)

4A. If a parenthetical phrase contains no point of access, do not attempt to add a backward-reading connective. The parentheses appear in the same place in all the permutations.

COPPER-ZIRCONUM <ALLOYS> (A LITERATURE SURVEY).

KOGNAK RIVER (WEST HALF)?, <DISTRICT OF <KEEWATIN> (EAST HALF)>.

4B. If there is an access point located within parentheses, insert a space plus left parenthesis as a backward-reading connective.

@LOWER <CRETACEOUS>? (<ALBIAN? (>)? OF THE <YUKON>.

AGE DETERMINATION AND GEOLOGICAL STUDIES (INCLUDING ISOTOPIC AGES). (Original Title)

S1 (S2 S3)S4 (Order of Elements in Original Title, where S2=the connective, and S4=the period or rest of the sentence following the parentheses)

AGE DETERMINATION? AND <GEOLOGICAL STUDIES? AND >? (INCLUDING <ISOTOPIC AGES - REPORT 3? (>). (Encoded Title)

S1? (S2 <S3? (>)S4 (Order of Elements in Encoded Title)

APPENDIX 1

IMPLEMENTATION REFERENCE SHEET

NEPHIS coding symbols are tagging characters which can be added by the human indexer to provide the computer with instructions for running the NEPHIS program and producing a printed or visually-displayed index.

"<" A "<" tags the beginning of a "nested phrase".

">" A ">" tags the end of a "nested phrase".

"< >" A pair of angular brackets defines a "nested phrase", consisting of one or more words.

OFFSHORE <EXPLORATION>. (Input String)

OFFSHORE EXPLORATION. (Permutations)
EXPLORATION. OFFSHORE -.

"@" This symbol indicates that the first word of an input string or a nested phrase is not desired as an access point in the finished index.

@INTRODUCTORY <MINERAL TABLE>. (Input String)

MINERAL TABLE. INTRODUCTORY -. (Permutation)

"?" The question mark suppresses coordinating expressions in certain of the entries while keeping them in others.

The "?" may tag a forward-reading connective:

ELECTRIC <POWER? IN <CANADA>>.

ELECTRIC POWER IN CANADA.
POWER IN CANADA. ELECTRIC -.
CANADA. POWER. ELECTRIC -.

Or the "?" may tag a backward-reading connective:

SAND? AND <GRAVEL? AND >.

SAND AND GRAVEL.
GRAVEL AND SAND.

Notes:
1. There must be the same number of left and right angular brackets in any input string.
2. Each input string must end in a period.
3. The beginning of an input string does not require a left angular bracket.

APPENDIX 2

IMPLEMENTATION PROJECT HISTORY

In late 1977 the NEPHIS indexing system had been implemented on only small databases of less than one hundred documents. At that time Craven proposed to test the wider applicability of NEPHIS by applying it to a well-defined set of at least three thousand Canadian government documents in the field of science and technology.

After receipt of NRC/NSERC Individual Operating Grant A3027 for 1978-1979, an appropriate set of documents was decided upon. These were the publications of the Department of Energy, Mines and Resources of the Canadian Federal Government as listed in Canadian Government Publications. Monthly Catalogue and the annual Catalogue. The time span chosen for this NEPHIS Implementation was from 1966 to April of 1978. For purposes of simplicity the one or two items in languages other than English were omitted.

The data elements recorded were basically the catalogue number as it appeared in the appropriate Catalogue and the document title coded with NEPHIS indexing characters. Subtitles were included when they occurred, but series titles were not. Periodicals were treated in the same way as series. Multi-volume works, not of a series or periodical nature and having separate volume catalogue numbers, were entered by using the full title with each volume.

M43-1/1969-1
GEOLOGY? AND <ECONOMIC <MINERALS>? AND >? OF <CANADA>. SECTION 1.

M43-1/1969-2
GEOLOGY? AND <ECONOMIC <MINERALS>? AND >? OF <CANADA>. SECTION 2.

The production of an index to this database of approximately three thousand documents was finished under Individual Operating Grant A3027 for 1979-1980.

APPENDIX 3

BIBLIOGRAPHIC REFERENCES

1. Craven, T. An experiment in teaching NEPHIS, a nested-phrase indexing system. Proceedings of the 4th Canadian conference on information science; 1976 May 11-14; London, Ontario. Ottawa: Canadian Association for Information Science/Association Canadienne des Sciences de l'Information; c1976: 131-139.

2. Craven, T. NEPHIS: a nested-phrase indexing system. Journal of the American Society for Information Science. 28(2): 107-114; 1977 Mar/Apr.

3. Craven, T. String indexing: NEPHIS. London, Ontario: School of Library and Information Science, The University of Western Ontario; 1977.

4. Craven, T. Linked phrase indexing. Information Processing and Management. 14(6): 469-476; 1978.

5. Craven, T. Need a subject index? Use NEPHIS. Creative Computing. 4(6): 94-95; 1978 Nov/Dec.

6. Craven, T. A government publication index using direct NEPHIS coding of titles. Journal of Information Science. 2(1): 13-21; 1980.

7. Farradane, J. Comparison of some computer-produced permuted alphabetical subject indexes. International Classification. 4(2): 94-102; 1977 November.

8. Svenonius, E. Current issues in the subject control of information. Library Quarterly. 47(3): 326-345; 1977 July.

FIGURE 1

NEPHIS INDEX SAMPLE PAGE

AAU/BNA PROJECT * 3 28-35
ABSTI * SEE Advisory Board on Scientific and Technological Information of National Research Council of Canada
ACCESS * SEE ALSO Confidentiality
ACCESS TO INFORMATION OF GOVERNMENT OF CANADA BY CITIZENS OF CANADA * 5 1-9
ADVISORY BOARD ON SCIENTIFIC AND TECHNOLOGICAL INFORMATION OF NATIONAL RESEARCH COUNCIL OF CANADA. Role in STI in Canada, 1969-1978 * 4 7-12
AGRICULTURE CANADA. Libraries. Network * 4 68-73
ANOMALOUS STATES OF KNOWLEDGE. Concept. Applicability to Retrieval of Data * 5 133-143
ARTIFICIAL INTELLIGENCE. Applicability to Communication between Computerized Systems & Human Beings * 2 120-128
ASK SERVICE * 1 35-42
ASSOCIATION OF ATLANTIC UNIVERSITIES * SEE ALSO AAU/BNA Project
ASSOCIATIONS IN INFORMATION SCIENCE IN CANADA. Federation. Desirability * 3 74-79
ASSOCIATIVE PROCESSING. Indexing * 5 93-101
ATLANTIC PROVINCES. Libraries. Use of Computerized Systems for Cataloguing * SEE ALSO AAU/BNA Project
ATLANTIC UNIVERSITIES COMPUTER STUDY * 1 31-34
AUDIO-VISUAL MATERIALS. Bibliographic Control. Role of National Film Board of Canada * 3 190-198
AUTOMATED * SEE ALSO Computerized
AUTOMATIC GENERATION OF THESAURI. Clustering of Terms. Use of Term Centroids * 4 124-136
AXES FOR CLUSTERING OF TERMS. Definition. Use of Factor Analysis * 5 183-197
B-N SOFTWARE RESEARCH. Experimental Use of Computerized System for Control of Information in Offices * 5 61-71
BIBLIOGRAPHIC CONTROL OF AUDIO-VISUAL MATERIALS. Role of National Film Board of Canada * 3 190-198
BIBLIOGRAPHIC CONTROL OF BOOKS IN FRENCH LANGUAGE RELATING TO HISTORY OF CANADA CITED IN NATIONAL UNION CATALOG PRE-1956 IMPRINTS. Role of WH&TC Project * 5 115-123
BIBLIOGRAPHIC CONTROL OF JURISPRUDENCE PERTINENT TO PROFESSIONAL LIABILITY INSURANCE IN QUEBEC * 4 74-78